Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-42-RC5, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "The use of personal weather station observation for improving precipitation estimation and interpolation" *by* András Bárdossy et al.

Marc Schleiss (Referee)

schleiss.marc@gmail.com

Received and published: 10 March 2020

Summary: The authors investigate the usefulness of personal weather stations (PWS) for quantitative precipitation estimation, with an emphasis on high-intensity rainfall events. A two-step procedure is proposed in which the first step is to select "good" PWS stations based on spatial consistency tests and the second step is to predict the rainfall rate at a target location using interpolation. The method is applied to a large dataset of professional and PWS data in Germany and the performance is evaluated using cross-validation.

General assessment: The topic is very relevant and the authors have some good ideas

for how to tackle the quality control and interpolation aspect of uncertain measurements from PWS. The 4 referees before me have already done a good job at pointing out relevant issues that need to be fixed/clarified before publication. Here are some additional ones:

Major comments:

a) The authors should provide more details about the kriging part. - How did you estimate the variograms? (with/without zeros?) - How do the variograms look like? - Which variogram model did you use and how well does it fit the empirical variogram? - How do you deal with cases in which there are not enough data to reliably fit a variogram? - How do you deal with spatial anisotropy and intermittency during interpolation?

b) Ordinary kriging makes rather strong assumptions about the data (such as second-order stationarity). The latter might not be very realistic in heavy localized rain events. Kriging is also relatively slow compared with other deterministic interpolation methods and its accuracy strongly depends on the density and number of primary observations. For example, the estimation and fitting of a variogram model (from a small number of samples) might introduce additional errors into your predictions that are due to modeling choices rather than the quality of the data. So my question is: why did you choose ordinary kriging? Please motivate this choice by some form of cost/benefit analysis, for example by comparing it to simpler, faster alternatives such as inverse weighted distance interpolation or bilinear interpolation (which make different modeling assumptions).

c) Related to the previous comment. Please note that during cross-validation, one part of the error is due to the spatial interpolation method that you use (i.e., kriging). If you had taken a different interpolation method (say IDW or Bilinear), perhaps the usefulness of the PWS data would have been different. I think it is important that you assess this part of the error by using at least one alternative non-parametric interpolation method other than kriging (e.g., bilinear interpolation). My point here is that in some cases, you might see improvement for one particular interpolation method but not for another.

d) The cross-validation part lacks crucial details about parameter estimation. For example, did you use the same variograms or recalculate them based on the selected subset of observations? Theoretically, you should recalculate the variograms on the smaller subset.

e) The second step (i.e., amount estimation) involves a quantile mapping. According to your Figure A1, this mapping is different for each PWS. However, this would mean that you need to estimate and fit a separate variogram model (with different nugget/range/sill) for each PWS location at which you want to interpolate. Is that correct? This would be computationally heavy. Please add more details to help me understand this.

f) Wind is known to cause localized biases in rain gauge measurements in the order of 10-30%. The latter are not stationary over time and space and can significantly affect the ordering of your data, therefore violating your model assumptions (i.e., monotonic link between quantiles of primary and secondary variables). This is not catastrophic but will occasionally affect the accuracy of your rainfall estimates and lower the reliability of your method. I think this issue should be clearly mentioned and discussed in the paper, along with the other limitations in the methodology mentioned by the other reviewers.

g) Tables 3 and 4: Your evaluation of the improvement in terms of a binary response (yes/no) is not very informative. Improved by how much? Some conditional error distributions (for both cases) might help shed some more light on best/worst case scenarios and what to expect in practice.

h) I agree with Lotte de Vos (referee 1) when she says that more details about the limitations of the method need to provided. I would go one step further and say that right now, the paper is heavily focused (biased?) towards demonstrating potential and improvements over the status quo. However, the numbers suggest there are also a lot

of cases in which the PWS data deteriorate the accuracy of the predictions. Perhaps you could show a few of these cases and comment on them. By explicitly showing what can go wrong, you may be able to provide concrete recommendations for future developments.

Minor comments:

1) Details about how correlation is calculated from binary variables are missing. I assume you used the Pearson linear correlation coefficient. But other choices are possible. Please clarify.

2) If a PWS gets accepted for a high value of alpha but not for a lower one, what does this mean?

3) Page 10, l.236: [...] indicating that the occurrence of precipitation can be well detected by the secondary network. This statement is misleading. According to previous studies, one of the most common type of errors in PWS are faulty zeros (i.e., the PWS gives zero rainfall during rainy conditions). The latter can occur without warning and only last for a few time steps (e.g., due to data transmission problems or temporary shielding of the gauge). The PWSs that pass the QC tests can still contain faulty zeros. Please reformulate to convey the right meaning.

4) Figure 5: How do you explain the red crosses that have higher correlation than the black crosses? Especially the ones with correlation coefficients close to 1. Are these computed from very small samples?

5) Appendix A, page 20, line 397: The equal weights of $\frac{1}{4}$ only applies to the case of an isotropic variogram.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-42, 2020.